

# 「主体と環境：身体性に関する一考察」 — 視覚移動ロボットにおける行動学習を例に —

浅田 稔

大阪大学工学部電子制御機械工学科 〒565 吹田市山田丘2-1

第2回創発シンポジウム講演資料(96/8/24)

## 1 はじめに

筆者の研究目標は、身体を有する物理的実体(ロボット)が環境、特に他者の存在によって、行動の複雑さを露呈する基本的内部構造の構築であり、結果として現れる行動の「観察者」の立場からの形容が「創発的」であったり、「知的」であったりすると考えている[1]。このことは、「創発」や「知能」を直接的かつ明示的に構築する事の矛盾や困難さを表していると考えられる。結論から先に述べると、「知的」行動を発現するためには、「身体」が必要であり、その意味は、

1. 環境の多様な構造をセンスできる感覚能力を有すること。
2. 環境に働きかける多様な運動能力を有すること。
3. 感覚と運動は不可分であり、密に結合していること。
4. 限られた資源(メモリなど)や処理能力の範囲でタスク(究極のタスクは、存続すること?)を遂行するために、感覚・運動空間を抽象化すること。
5. 抽象化は、身体への基本埋め込みと経験(環境との相互作用)に依存するので、抽象結果は、身体を有する主体中心の主観的表現であり、その評価は行動結果でなされる。

と考えている。設計原理としては、以下の2点に絞られる。

- 上記の身体性を有するエージェントの内部構造の設計。
- 内部構造を発達させるタスクと環境の与え方。

以下では、まず、人工知能やロボティクスの基本問題である世界の分節化に対する一手法を示し、その拡張をベースとして、身体性を有する主体が「知的」行動を発

現できる環境の構造、特に他者の存在と「知能」との関わりについて論ずる。

## 2 強化学習とサッカーロボット

基本問題を明示するために、我々が最初にやった視覚に基づくサッカーロボット<sup>1</sup>研究を例に考える[2]。視覚情報から得られるのは、ボールとゴールの画像情報だけなので、それらの特徴を反映した状態として、ボールの位置と大きさ、ゴールの位置と大きさ、及び向きについて図1に示す状態空間を用いた。

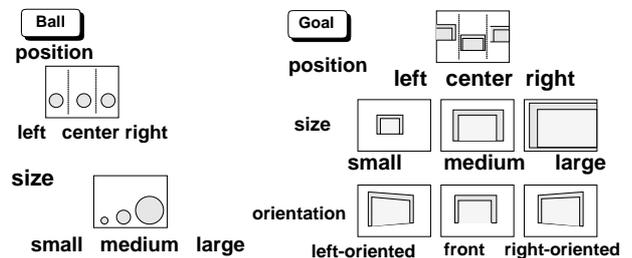


図1: ボールとゴールの状態空間

行動の方は、ブルトーザの二つの独立したモータへの前進、停止、後退命令の組合せで、ストップを除いた8種類の行動からなる。尚、これらの物理的意味などは、ロボットが知る由もない(ことにした)。最初の問題は、実際のセンサーやアクチュエータを反映した状態と行動は、必ずしも同期しないことである。図2に示すように、ゴールから遠く離れた所にいる場合、前進行動を一回とっても状態は変わらない。ところが、中間位との堺目だと1回

<sup>1</sup>最初の成果を論文発表した時、サッカーロボットではないと酷評された記憶がある。シューティングロボットという、銃で撃つ感じがしたのと、これからもっと多くのロボットを入れ込んでチームプレイをやるんだという意気込みというか、プレッシャーをかけるために、あえて「サッカーロボット」と呼ばせて頂く。但し、ジャーナルペーパーなどで、タイトルにサッカーロボットを題したものはなく、...

で、状態が変わってしまう。つまり、同じ状態で、同じ行動をとっても状態遷移のバラツキが大きくなり、学習が進まない。そこで、ここでは、状態が変わるまで、同じ動作を続け、その一連の動作を行動とした。これは、以下で述べるその他の例でも、基本原理として用いている。

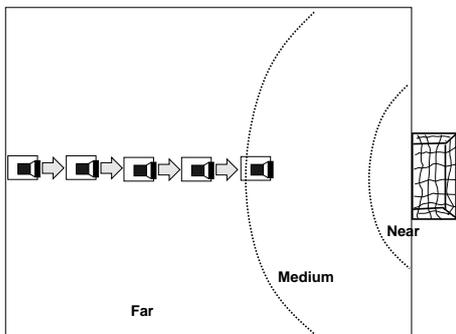


図 2: 状態と行動のズレ問題

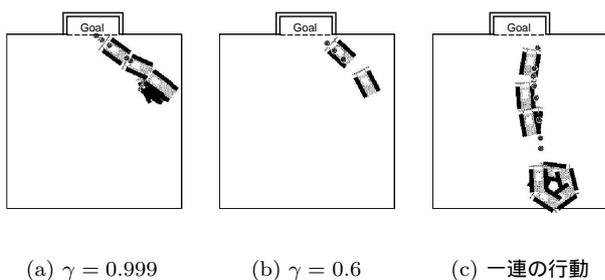


図 3: シミュレーション中の行動

シミュレーション中のロボットの行動例を示そう。Q-学習における行動価値の更新式中の減衰係数 ( $\gamma$ ) によって、表出するロボットの行動が変わってくる。Fig.3(a)では、 $\gamma$ の値が大きいので、どんなに時間を掛けてもゴール時の報酬が、あまり変わらない。そのため少しでも確実にシュートしようと、よりよい位置に移動してからシュートしているのに対し、(b)では、 $\gamma$ の値が小さく、早くゴールしないと報酬がもらえないので即座にシュートしている。尚、(a,b)では、比較のためにスタート地点は同じである。(c)では、学習した政策を用いた一連の行動を表す例を示した。最初にボールを見失い、発見し、ドリブルして、最後にシュートしている。シミュレーションはうまくいっている。実機でもそれなりに成功した。問題は、状態行動空間を与えた時点で、答えを教えたようなものだと感じられ、状態と行動のズレよりも、うまく行動できる状態・行動空間をいかに構成するかが、より基本問題であることが分かった。

## 2.1 状態空間の自律的構成

状態空間を構成する問題は、基本的に重要で非常に困難な問題である。この例では、すでにボール(赤領域)、ゴール(青領域)が抽出されているから、かなりやさしくなっているかもしれない。それらの領域特徴を状態空間の次元として採用したが、一応、これも自動抽出を試みた。多数の動作から得られる画像情報から面積、重心、モーメントなどの基本特徴を並べ、主成分解析で主成分を求めると、ほぼ自明だが、ボールやゴールの位置、大きさ、傾きなどに相当する特徴が対応した。次の問題は、この軸をどのように分けるかの「分節化」の問題である。先に述べたように、基本原理(「状態が変わるまで、同じ動作を続け、その一連の動作を行動とした。」)から、逆にゴール状態に到達するまでの動作が変化しないならば、それらの見掛けが異なっても、同じ状態に区分する方針をとり、初期状態としてゴール状態とその他の二つにしか分かれていなかった状態空間を再帰的に分節化していった[3]。図4と表1に比較結果を示す。図では、ボールの大きさとゴールの大きさの2次元に投影している。格子状の箱は、先の研究で人間が与えた状態分割で、楕円状に投影されている領域が自律的分節した結果である。形状が大きく異なること、また大きさも異なり、探索時間が激減し、シュートの成功率も上がっている。

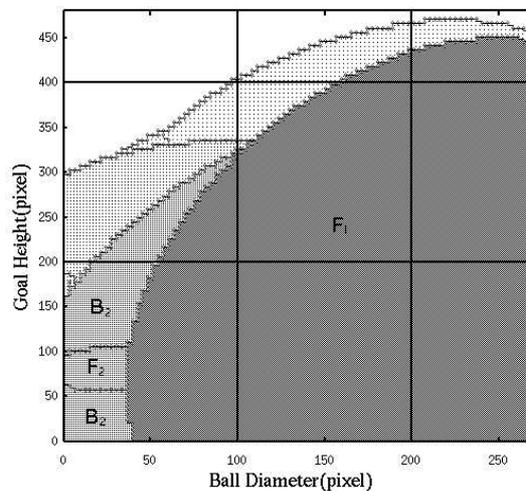


図 4: 状態の自律的分節化

この例では、以下の二つが重要と考えている。

1. 人間が与えた状態空間はロボットにとって最適とは限らず、ロボットが自らの経験から自分が見切るべき状態を決定すべきで、世界の分節化問題と同じであると考える。すなわち、主体(感覚と行動能力の規

### 3 他者の存在と「知能・創発」

表 1: 人間が与えた状態空間と自律的に構成された空間

	状態数	探索時間 (1/30 秒)	成功率 (%)
人間が与えた	243	500M*	77.4
自律的に構成	33	41M	83.3

\* Q 学習の時間を指す。

定)と環境に依存して、世界の見方が変わる。

2. 本来、状態と行動空間の構成問題は、「鶏と卵」問題に類似し、相互に密接に関連している。何らかの最低限の拘束を導入しないと、相互に規定できない。ここでは、行動の最小単位を規定し、状態が変化するまで、同じモーターコマンドを続けることにより、続ける長さをパラメータとして行動空間と状態空間を同時に構成していると考えられることも可能である。

後者は、プリミティブなモーターコマンドを拘束として、行動空間を時間軸に抽象化したと考えられるが、物理的な自由度が2で、行動空間の空間的な抽象化の対象としては自由度が少なく、その結果、抽象化が知覚中心となっている。

人間は、モノを掴む時、近くにあれば、すぐに手で、ちょっと離れていれば手を伸ばして、もっと離れていれば立上り、移動してモノを掴む。このとき、利用している物理的自由度はかなり多いにも関わらず(実際、そのような超多自由度のロボットの制御は非常に困難な問題)、人間は難なくこなしているように見える。これは、例えば、実際の手の関節数に関わらず、ものを掴む時は、開閉の1自由度で十分であるが、このような運動空間の抽象化は、いかにして獲得されるであろうか? 視覚を始めとする感覚情報の抽象化(分節化)は、主体の運動能力に依存すると述べたが、逆に、運動能力の抽象化(分節化)も、主体の感覚能力に依存する。このための、運動能力に対する、最低限必要な規定(埋め込み)は何か、重要な問題となってくる。

「身体性」の本質は、主体の「多様な感覚情報」、「多自由度の運動能力」を、物理的実体として限られた処理能力の範囲で、タスクを遂行(存続することが究極のタスク?)するために「抽象化」することであり、その結果として「知的」振舞が可能になると考えている。環境構造の複雑さ、特に「他者の存在」が、抽象化のレベル向上を必然的に伴うことを、視覚を例に、以下で説明を試みる。

知能の本質に関する議論として、社会性、他者の存在、模倣などがキーワードとして挙げられている [4, 5]。特に大澤は、自己を否定し得る可能性のあるものやことを「他者性」とし、未来性も含めており、時間との関連性を印象づける。離人症の例 [6] から、自己の存在を時間と直接関連づける木村 [7] の説との関連性もあると考える。

知能のレベルが、環境との相互作用の複雑性にあると考えると、視覚を基にした環境との相互作用をベースとする我々の研究でも、他者の存在が、複雑さを増すであろうことや、自己の規定などに関する議論が可能かと思うので、以下に試みる。

1. 自己の規定：運動機能を有する主体が、運動指令と視覚情報との直接的な相関がとれる場合、すなわち、運動指令を発した自己の運動結果が視覚で確認され、大澤の言を借りれば、運動指令を出した自己を裏切らない範囲が自己と規定し得る範囲。例えば、自分の腕や指は、他人のそれとどうやって区別するか? 視覚に限らず、あらゆるセンサ情報が、自己のモーターコマンドとある種の相関を持つ可能性がある。

細田、浅田による汎用視覚サーボの手法 [8] では、手先の運動を視覚で捉え、画像ヤコビアンをオンライン推定し制御しているが、推定可能という意味で自己の部分と判断可能である。

2. 静止環境：自己の運動との直接的な相関により、自己と規定された範囲を除いて、切り出し可能な対象。切り出せれば制御可能。Nakamura and Asada は、オプティカルフローによる行動学習で「運動スケッチ」 [9] と呼ばれる視覚移動ロボットの内部表現を提案しており、その初期段階では、静止環境に対するフローパターンとモーターコマンドの相関により運動空間を抽象化した。逆にいえば、そのようなモーターコマンドとの直接的相関がとれる部分が静止環境とみなせる<sup>2</sup>。
3. 受動主体：自己や他者の運動の帰結により、静止環境内で運動したり、停止したりするもの。自らは状態変化を起こさないもの。サッカーロボット [2] では、ボールが相当する。状態空間の自律的構成 [3] では、静止環境(ゴール)と受動主体(ボール)を同時に含む空間なので、静止環境のみを主体とするゴール探索

<sup>2</sup>1, 2ともにロボットの運動指令と視覚情報の直接的相関であるが、1が観測系を除く身体への運動指令に対し、2が観測系の運動指令を暗に示していることに注意。同時に呈示されたときでも理論的には識別可能であるが、計算は複雑になると予想される。

より、複雑度が高いと考える。オンラインで状態分割を行った高橋らの研究[10]では、ロボットがボールにたどり着くまでは、ボールは静止環境として扱われ、たどり着いてからは、位置、サイズが固定と表現される。よって、線形モデル(後者は勾配ゼロ)で近似でき、空間的な分割だけで対応可能であった。但し、以下の他者(ゴールキーパーなど)が入ると、現状では対応が困難である。

4. 能動主体:直接的には、自己の運動との単純な相関を持たない。大澤の意味では、自己を否定する可能性のあるもの。視覚情報からの相関では、単純な相関を持たないものとして、最初、排他的に扱われるが、自己の運動との高度な相関(協調、競争、欺瞞)により、規定される対象。複雑度は一挙に増す。サッカーロボットの例では、ゴールキーパーを避けてシュートしようとするロボットの研究で、他者との駆け引き(回避)行動と、シュート行動を個別に学習し、その後、再学習により統合したが[11]、明示的には、他者の意識はロボット主体にはない。ただ、行動選択の複雑さが増していることは確かであろう。

以上述べてきたように、環境との相互作用の複雑さが、エージェントの内部構造の発達を促し、行為の多様性を生み出すと考えられる。現状では、我々はこれらを実現できる統一的な基本内部構造を提案していないが、能動主体に対する行動理解の一手法を行っている[12]ので、以下に簡単に説明する。

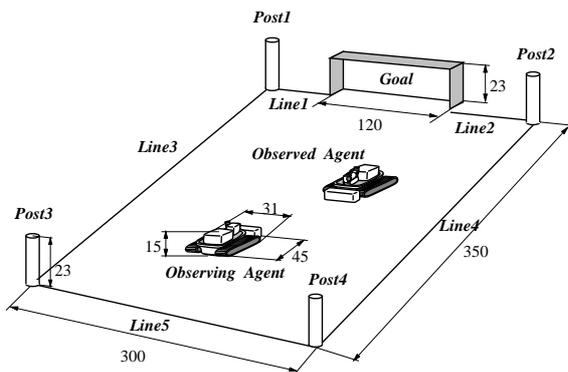


図 5: 環境

設定として、図5のような環境(ゴール、ポスト4本、ライン5本)を考える。観測者以外に、もう一体のエージェントが存在し、そのエージェントの行動を理解することを目的とする。エージェントは同一のものを使用し、

PWS(Power Wheeled Steering)を採用する。また、自身の幾何学的パラメータや動的特性などに関する知識はもっていないと仮定する。エージェントにはカメラが搭載されており、このカメラだけから、環境の情報を獲得しなければならない。具体的には、画像上での位置の  $x$  座標、大きさなどを検出する。エージェントに関しては、向きも検出できると仮定する。

観測されるエージェントの行動戦略は、(a)「環境内に存在する物体だけに依存」、(b)「自身の内部状態だけに依存」の2種類あり、実際の戦略はこの2種類の戦略の中間をとっていると考えられる。そこで観測される行動戦略として、(i) 静止している、(ii) ゴールに向かう、(iii) 観測エージェントに向かう、(iv) ランダムウォークの4種類を準備する。実際の行動は、環境からの制約を受けるため、必ずしも「自分自身の内部状態だけに依存」するわけではないが、(iv) は (b) の一例と考えられる。また、(ii),(iii) は (a) のなかで、対象物が静的か動的かの違いがある。(i) は他のエージェントが単純に静止環境の一部とみなせる例である。ここで、一試行の観測中に、観測されるエージェントの行動戦略は、上記のいずれか一つだけを使用しているものとし、途中で変更しないと仮定する。

他のエージェントの行動(状態遷移)を理解するためには、そのエージェントを観察しなければならない。しかし、観測するエージェントがランダムに行動していたのでは、効率が悪い。つまり、観察のための行動戦略として、相手エージェントを注視し続けるような戦略が必要となり、ここでは単純な制御則によりこれを実現する。

具体的には、多変量データ解析を用いて観測データを主成分解析する。相手の行動を説明するための変数(説明変数)の候補として、それぞれ

- 自分の行動:2種類(前進速度  $v$  と角速度  $\omega$ )
- 相手エージェント:3種類(画像上での位置の  $X$  座標、大きさ、向き)
- ゴール:3種類(相手エージェントと同様)
- ポスト:2種類(画像上での位置の  $X$  座標と大きさ)
- ライン:2種類(画像上での位置の  $Y$  座標と傾き)

を準備し、この組合せでデータ行列の説明変数は決定される。例えば、相手エージェントとゴールだけが見えている場合には、説明変数の個数は自分の行動も含めて合計  $2 + 3 + 3 = 9$  個となる。

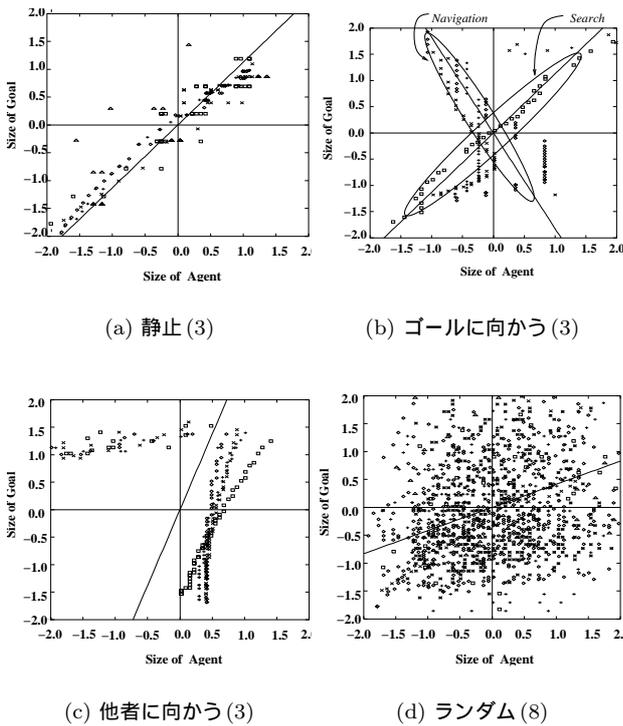


図 6: シミュレーション結果

図 6の (a) ~ (d) に、それぞれの行動について獲得された、相手エージェントとゴールに関するデータ行列を列標準化したものの一部を示す。データの可視化のため、 $X$  軸を画像上でのエージェントの大きさ、 $Y$  軸を画像上でのゴールの大きさとする。括弧内の数字は、累積寄与率 90% 以上の主成分の個数である。

(d) の場合には、画像の解像度の低さのため、標準化したデータの分布には、疎密が見られるが、ほぼ様に観測データが分布している。他のそれぞれの行動には、

1. (a) 相手が大きくなると、ゴールも大きくなる。
2. (b),(c) 相手の大きさは一定であり、ゴールの大きさは変化する。

といった特徴がある。(b),(c) は、エージェントの向きの軸により特徴づけることができる。主成分分析により、これらのデータを説明できるような、新しい軸(主成分)が発見できる。このような関係が、ポストやラインなどに対しても求められる。

次に、実際にゴールへ向かう行動戦略を主成分を用いて判別した結果を図 7 に示す。ここで、 $Y$  軸は情報損失量を表す。相手が静止している場合には、主成分がかな

り異なるため、結果として情報損失量が他と比べて、非常に大きな値となり、図 7 には示していない。観測の初期段階で、自分自身に向かう行動と識別を失敗しているが、だいたい場所で正確に識別できた。

失敗の原因として、ゴールへ向かう行動戦略が複数の戦略から構成されていることが挙げられる。ゴールへ向かう行動戦略は 1. ゴールの探索、2. ゴールまでのナビゲーションの 2 つの戦略に分解できる(図 6(b) 参照)。獲得した主成分は、2. の方の戦略を良く説明しているのに対し、観測の初期段階では、1. の行動戦略をとっている。結果として、この間の情報損失量は増加することになる。

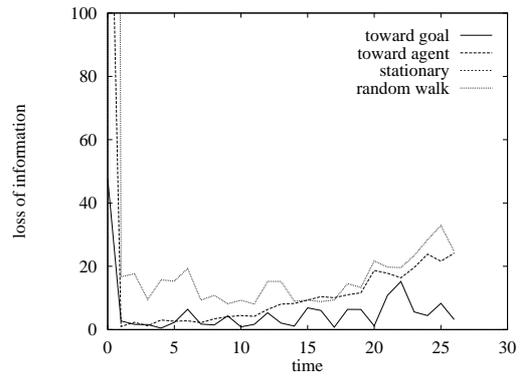


図 7: 識別結果

このように、他者の簡単な行動戦略は識別可能であるが、本質的な問題として部分観測 (Partially Observable) による限界を免れていない。つまり完全な環境モデルや相手の観測モデルを想定できない。これを完全には克服できないが、観測者が漫然と観測するだけでなく、識別のための情報獲得の行動を誘発できるメカや行動の構造化が望まれる。

#### 4 発達と導き

欺瞞が自己とか自意識と関連することが、動物行動学 [13] などでも論議されているので、学習、発達過程で「フェイント」できるサッカーロボットができないものかと考えている。但し、発達過程を導く何らかの教示か模倣が必要であると考えられる。それは、言語獲得時の母親であり [14]、直接指示の教示 [5] であろう。サッカーロボットでも、相手がいない環境ですら、ボールをゴールにシュートさせるために、ボールをゴールの近くに設置しシュートしやすい環境から始めた。またゴールキーパーがいる場合でも、ゴールキーパーが最初から上手だと、プレイヤーは全くシュート出来ず、自信を持ってない。そこで、やさしい

タスクから学習する規範 (Learning from Easy Mission: LEM)[15] に則り、最初ゴールキーパーを静止させ、徐々に速度を挙げることで、最初から同じ速度で動くキーパーを相手にした場合と比較して格段の学習時間の短縮が可能になった [16] . 図 8 に、その効果を示そう . 最初にゴールキーパーを静止させ、次に学習ロボットの最大速度の半分、最後に同じ速度まであげた場合のシュートの成功率が実線、最初から同じ速度を持たせた場合が破線で示されている .

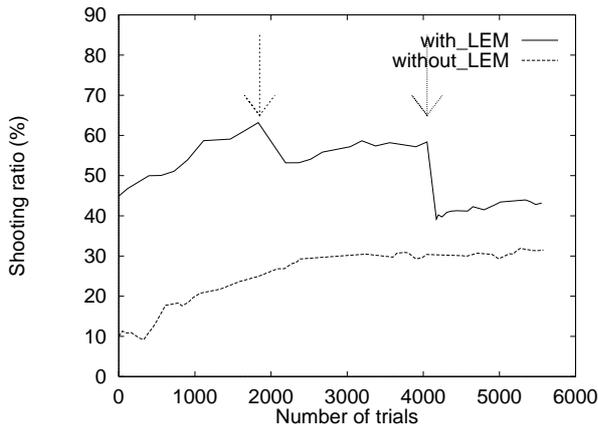


図 8: やさしいタスクからの学習による効率化

## 5 おわりに

環境との相互作用の複雑さを反映するエージェントの内部構造の設計の重要性、及び内部構造を発達させるためのタスクや環境の提供の仕方の重要性について、サッカーロボットを例に述べた . サッカーロボットは、限られた環境やタスクではあるが、基本問題を含んでおり、AI やロボティクスの標準問題として十分その価値があると感じている [17] .

基本構造の設計や環境呈示がうまく (?) 運ばば、今後、より多くのエージェントの存在により、より高次の相互作用が必然的に生じ、これにより行動の多様性が発現されると期待される .

最後に、本稿に関連する討論に加わった、本学細田耕博士、博士課程の内部英治君、高橋泰岳くん、ポスドクの中村恭之君に感謝する .

## 参考文献

[1] D. McFarland and T. Bösser. *Intelligent Behavior in Animals and Robots*. MIT Press, 1993.

- [2] 浅田, 野田, 依積田, 細田. “視覚に基づく強化学習によるロボットの行動獲得”. 日本ロボット学会誌, Vol. 13:1, pp. 68–74, 1995.
- [3] M. Asada, S. Noda, and K. Hosoda. Non-Physical Intervention in Robot Learning Based on LfE Method. In *Proc. of Machine Learning Conferen Workshop on Learning from Examples vs. Programming by Demonstration*, pp. 25–31, 1995.
- [4] 大澤. 「知能の社会性」. 日本ロボット学会誌, Vol. 14, No. 4, pp. 15–19, 1996.
- [5] 國吉. 「実世界エージェントにおける注意と視点-情報の分節・統合・共有-」. 人工知能学会誌, Vol. 10, No. 4, pp. 507–514, 1995.
- [6] 中島. 「手足を持った人工知能」. 日本ロボット学会誌, Vol. 14, No. 4, pp. 12–14, 1996.
- [7] 木村敏. 「時間と自己」. 中公新書, 1982.
- [8] 細田, 浅田. 「構造やパラメータに関する先験的な知識を必要としないフィードフォワード補償器を持つ適応型ビジュアルサーボ系の構成」. 日本ロボット学会誌, Vol. 14, No. 2, pp. 159–165, 1996.
- [9] T. Nakamura and M. Asada. Motion Sketch: Acquisition of Visual Motion Guided Behaviors. In *Proc. of IJCAI-95*, pp. 126–132, 1995.
- [10] 高橋, 浅田, 細田. 状態空間の自律的分割による実ロボットの長時間学習. ロボティクス・メカトロニクス講演会 96 予稿集, pp. 121–124, 1996.
- [11] M. Asada, E. Uchibe, S. Noda, S. Tawaratsumida, and K. Hosoda. “Coordination Of Multiple Behaviors Acquired By Vision-Based Reinforcement Learning”. In *Proc. of IROS94*, pp. 917–924, 1994.
- [12] 内部, 浅田, 細田. 他のエージェントの行動理解 — サッカーロボットにおける強化学習のマルチエージェント環境への適用に向けて —. 第 13 回日本ロボット学会学術講演会予稿集, pp. 241–242, 1995.
- [13] リーキー著, 馬場悠男訳. 「ヒトはいつから人間になったか」. 草思社, 1996.
- [14] 正高. 「身体運動は言語獲得にどのような役割を果たすか」. 日本ロボット学会誌, Vol. 14, No. 4, pp. 31–34, 1996.
- [15] M. Asada, S. Noda, S. Tawaratsumida, and K. Hosoda. Vision-Based Reinforcement Learning for Purposive Behavior Acquisition. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 146–153, 1995.
- [16] M. Asada, E. Uchibe, and K. Hosoda. Agents That Learn from Other Competitive Agents. In *Proc. of Machine Learning Conferen Workshop on Agents That Learn from Other Agents*, pp. 1–7, 1995.
- [17] 北野, 大沢, 松原, 野田, 國吉, 浅田, 稲葉. 「特集: サッカーエージェント – RoboCup –」. *bit*, Vol. 28, No. 5, pp. 22–52, 1996.